

Errors Corrected by the STIC Systems Branch

Serial Number: 08/487,032A

CRF Processing Date:

Edited by:

Verified by:

180511
3/24/95
PP

Changed a file from non-ASCII to ASCII

Changed the margins in cases where the sequence text was "wrapped" down to the next line.

Edited a format error in the Current Application Data section, specifically:

Edited the Current Application Data section with the actual current number. The number inputted by the applicant was the prior application data; or other _____

Added the mandatory heading and subheadings for "Current Application Data".

Edited the "Number of Sequences" field. The applicant spelled out a number instead of using an integer.

Changed the spelling of a mandatory field (the headings or subheadings), specifically:

Corrected the SEQ ID NO when obviously incorrect. The sequence numbers that were edited were:

Inserted or corrected a nucleic number at the end of a nucleic line. SEQ ID NO's edited:

Corrected subheading placement. All responses must be on the same line as each subheading. If the applicant placed a response below the subheading, this was moved to its appropriate place.

Inserted colons after headings/subheadings. Headings edited included:

Deleted extra, invalid, headings used by an applicant, specifically:

Deleted: non-ASCII "garbage" at the beginning/end of files; secretary initials/filename at end of file; page numbers throughout text; other invalid text, such as _____

Inserted mandatory headings, specifically:

Corrected an obvious error in the response, specifically:

Edited identifiers where upper case is used but lower case is required, or vice versa.

Corrected an error in the Number of Sequences field, specifically:

A "Hard Page Break" code was inserted by the applicant. All occurrences had to be deleted.

Deleted *ending* stop codon in amino acid sequences and adjusted the "(A)Length:" field accordingly (error due to a PatentIn bug). Sequences corrected: _____

Other:

*Examiner: The above corrections must be communicated to the applicant in the first Office Action. DO NOT send a copy of this form.

3/1/95

PAGE: 1

RAW SEQUENCE LISTING
PATENT APPLICATION US/08/487,032A

DATE: 03/25/97

TIME: 10:10:00

INPUT SET: S16452.raw

This Raw Listing contains the General
Information Section and up to the first 5 pages.

ENTERED

1 SEQUENCE LISTING
2
3 (1) General Information:
4
5 (i) APPLICANT: DOUGLAS SMITH
6
7 (ii) TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES
8 RELATING TO HELICOBACTER PYLORI FOR
9 DIAGNOSTICS AND THERAPEUTICS
10
11 (iii) NUMBER OF SEQUENCES: 880
12
13 (iv) CORRESPONDENCE ADDRESS:
14 (A) ADDRESSEE: LAHIVE & COCKFIELD
15 (B) STREET: 60 State Street
16 (C) CITY: Boston
17 (D) STATE: Massachusetts
18 (E) COUNTRY: USA
19 (F) ZIP: 02109-1875
20
21 (v) COMPUTER READABLE FORM:
22 (A) MEDIUM TYPE: Floppy disk
23 (B) COMPUTER: IBM PC compatible
24 (C) OPERATING SYSTEM: PC-DOS/MS-DOS
25 (D) SOFTWARE: PatentIn Release #1.0, Version #1.25
26
27 (vi) CURRENT APPLICATION DATA:
28 (A) APPLICATION NUMBER: US 08/487,032
29 (B) FILING DATE: 07-JUNE-1995
30
31 (viii) ATTORNEY/AGENT INFORMATION:
32 (A) NAME: Mandragouras, Amy E.
33 (B) REGISTRATION NUMBER: 36,207
34 (C) REFERENCE/DOCKET NUMBER: GTN-001
35
36 (ix) TELECOMMUNICATION INFORMATION:
37 (A) TELEPHONE: (617)227-7400
38 (B) TELEFAX: (617)227-5941
39
40
41 (2) INFORMATION FOR SEQ ID NO:1:
42
43 (i) SEQUENCE CHARACTERISTICS:
44 (A) LENGTH: 1527 base pairs
45 (B) TYPE: nucleic acid
46 (C) STRANDEDNESS: double

INPUT SET: S16452.raw

47 (D) TOPOLOGY: circular
48
49 (ii) MOLECULE TYPE: DNA (genomic)
50
51 (iii) HYPOTHETICAL: NO
52
53 (iv) ANTI-SENSE: NO
54
55 (vi) ORIGINAL SOURCE:
56 (A) ORGANISM: Helicobacter pylori
57
58 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:
59
60 ATGTGTTCTC AGGAAATTTT ATCAAGCTTG CAAACCATT A TTGCCGAACA ATTTTCTATA 60
61 AATATCATCA CTCAGCTTGC TAATAAACTC ACACAAGTTA AAAATCTAAA TTTTTTGAG 120
62
63 AATAAAGACC ATACTATCAA GCTTAACACT ATCCATAACG GACTGCACAT CCGCCCCCTA 180
64
65 AATTATGTCA GTAATCTTTT TTTCAATCTA CAACGCATTA TAGGGCTTAT CAGTCTGTTT 240
66
67 GGGATATTAT TTTCCATTAG TATTATCTA CCCTTTATAA TGATTTTGCG AACAGTGCCT 300
68
69 TGTATTCTCA TTTCCAACCA TATAGCAAAA AAACATAGTG CTTCCATAGA TAAACTTCAA 360
70
71 GACCAAAAAG AAAGCATGCA AAATTACTTA TACTCTGGAC TAGATAACCA AAAGAACAAAG 420
72
73 GACAACCTAT TATTTAACCTT CATGCTAAAT TTTCACCATA AATTTATTGA AACAAAAGAA 480
74
75 TTGTATCTCA ATAATTTGT GAAAGTAGCC CAAAAAAACT TAATATTTAC CATATATGCT 540
76
77 GATGTTTAA TCACCACTCT AAGTATTGCA CTATTTTTC TAATGGTTT TATTATCCTT 600
78
79 TCAAAATTAA TTGGTGTGGG AGCAATTGCT GGGTATATCC AAGCATTAG CTCTACCCAA 660
80
81 CAACAACTAC AAGATTTATC ATTTTATGGA AAGTGGTTTT TTGCTATCAA TAAATACCTT 720
82
83 GAAAATTATT TCTGTATTT AGATTACAAA ATACCGAAAC CAGAAACACA AATCAAATTAA 780
84
85 GAAGAAAAAA TCCATAGCAT TACATTTGAA AATATTAGTT TCTCTTATCC TAATTCAAAA 840
86
87 CTTATTTTG AAAACTTAA TCTCTTTA CACTCTAATA AAATTTATGC ATTAGTCGGC 900
88
89 AAGAATGCTA GCGGAAAAAG CACGCTGATT AATTTATTAT TAGGTTTTA TACCCCAAAT 960
90
91 TCAGGTCAAA TTATCATTAA TAACAAATAC CCATTACAAG ACTTGGAACT AAATAGCTAC 1020
92
93 CATCAACAAA TGAGTGCCAT ATTTCAAGAT TTTTCTCTTT ATGCTGGTA TAGCATTGAT 1080
94
95 GATAATCTT TTATGCAAAA CAATATCACT AAAGAGCAAT TGAAGCAAAA AAGAGAAATA 1140
96
97 CTAAAATCTT TTGATGAGAA TTTCAAAAT TGTCTTAATG ATTGCAACAA CACACTATTT 1200
98
99

INPUT SET: S16452.raw

100 GGAGCGCAAT ATAATGGGGT AGATTTTCT TTAGGTCAAA AGCAACGCAT AGCTACCATG 1260
101
102 AGAGCCTTT TAAAACCAAG TAATTGCATT GTTTAGATG AGCCAAGCAG CGCCATCGAT 1320
103
104 CCCATTATGG AAAAAGAGTT TTTAGATTT ATTTTTAAAA AATCGCAATC TAAGATGGCT 1380
105
106 TTAATTATTA CACACCGCAT GAATAGTGTCA AAGCAAGCTA ATGAAATTAT CGTGTAGAT 1440
107
108 CAAGGCAAAC TAATAGAACCA GGGCAACTTT GAAACCTTA TGAAAAAACAA GGGATTATTT 1500
109
110 TGCGAATTGT TTTTGAAACA ACAATAC 1527
111
112
113 (2) INFORMATION FOR SEQ ID NO:2:
114
115 (i) SEQUENCE CHARACTERISTICS:
116 (A) LENGTH: 399 base pairs
117 (B) TYPE: nucleic acid
118 (C) STRANDEDNESS: double
119 (D) TOPOLOGY: circular
120
121 (ii) MOLECULE TYPE: DNA (genomic)
122
123 (iii) HYPOTHETICAL: NO
124
125 (iv) ANTI-SENSE: NO
126
127 (vi) ORIGINAL SOURCE:
128 (A) ORGANISM: Helicobacter pylori
129
130 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:
131
132 ATGGAGCGCA AGACGCTCCA GAGCATTAA TGTTAATAA AAAAAGAAAT GATGAGACCA 60
133
134 AAAGGTATTC TAATGAATTG TTGCAGGSCT TGGAAACACC AGGTTCTTAA GCAAAGCACG 120
135
136 ACAGGTTTAG TGGTGTGAG CATTATCTCT TCTACAGCCC CCTTTATTGG TTTGTTGGG 180
137
138 ACGGTAGTTG AAATTTAGA AGCGTTAAC AATTTGGCG CGTTAGGTCA AGCTTCTTTT 240
139
140 GGAGTGATCG CACCCATTAT TTCTAAGGCG CTTATGCCA CCGCTGCAGG GATTTAGCA 300
141
142 GCCATTCCAG CCTATTCTTT TTACTTGATC TTAAAGCGCA AGGTGTATGA TTTATCGGTT 360
143
144 TATGTGCAGA TGCAAGTGGAA TATTTGTCT TCTAAAAAA 399
145
146
147 (2) INFORMATION FOR SEQ ID NO:3:
148
149 (i) SEQUENCE CHARACTERISTICS:
150 (A) LENGTH: 474 base pairs
151 (B) TYPE: nucleic acid
152 (C) STRANDEDNESS: double

INPUT SET: SI6452.raw

153 (D) TOPOLOGY: circular
154 (ii) MOLECULE TYPE: DNA (genomic)
156 (iii) HYPOTHETICAL: NO
158 (iv) ANTI-SENSE: NO
160 (vi) ORIGINAL SOURCE:
162 (A) ORGANISM: Helicobacter pylori
163 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:
165
166 ATGCATGAAC GCATTGAAAG AGGTATTRGA AATAATGAAT GTAAAGAAAT TTTTGGCAAT 60
167
168 GAACTCAAAC AAAGAAAGAC AAAATTGATT GAAGACATAG AACGGCGGTT CAAAGAATGC 120
169
170 GAGGAACAAT TCCGTGGAAG TGTAGGAAAA AATATTGAAC AACTTGAAGA AAGAGTTAAA 180
171
172 GATTCTCTAG CGATTATAAA ACGCATCAAT AACCTTGGTC TTAATCCTAA TTCTAATTTT 240
173
174 AATATGGATA GCGGCATTGA TACAATAGGC TTATTTAGTT CAATAGGAGG TTTGGTGTG 300
175
176 CTTCTATTGA CGCCTGTAGT AGGTGAGTTT GCGTTAATTG CAGGAGTGGG TTTAGCATT 360
177
178 GTGGGGTAG GTAAATCAAT ATGGAGTTTT TTTGATTCAAG ATTATAAAAAA ATCCCAACAA 420
179
180 AGAAAAAGAAG TGGATAAGAA TTTACATCAA ATTTGCGAAA AATTGTGCAG GATG 474
181
182
183 (2) INFORMATION FOR SEQ ID NO:4:
184
185 (i) SEQUENCE CHARACTERISTICS:
186 (A) LENGTH: 336 base pairs
187 (B) TYPE: nucleic acid
188 (C) STRANDEDNESS: double
189 (D) TOPOLOGY: circular
190
191 (ii) MOLECULE TYPE: DNA (genomic)
192
193 (iii) HYPOTHETICAL: NO
194
195 (iv) ANTI-SENSE: NO
196
197 (vi) ORIGINAL SOURCE:
198 (A) ORGANISM: Helicobacter pylori
199
200 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:
201
202 ATGCCTGGCG TGTATCAAAT GAGTATAGAG CCTCTTTAA AAGAATGCGA AGAATTAGTG 60
203
204 GGTTTAGGCA TAAAAGCCGT TTTATTGTTT GGCATTCTA AACATAAGGA CGCTACAGGA 120
205

INPUT SET: S16452.raw

206 AGCCATGCGT TAAATAAGGA TCACATTGTC GCAAAAGCTA CGAGAGAAAT TAAAAAACGA 180
207
208 TTTAAGGATT TGATCGTTAT AGCGGATTTG TGTTTTGCG AATACACCGA CCATGGGCAT 240
209
210 TGCAGGGATTT TAGAAAACGC TTCTGTGTCT AACGATAAAA CGCTAAAGAT TTTAAATCTT 300
211
212 CAAGGGCTTA TTTTGCTGAA AGCGGTGTGG ATATTC 336
213
214
215 (2) INFORMATION FOR SEQ ID NO:5:
216
217 (i) SEQUENCE CHARACTERISTICS:
218 (A) LENGTH: 195 base pairs
219 (B) TYPE: nucleic acid
220 (C) STRANDEDNESS: double
221 (D) TOPOLOGY: circular
222
223 (ii) MOLECULE TYPE: DNA (genomic)
224
225 (iii) HYPOTHETICAL: NO
226
227 (iv) ANTI-SENSE: NO
228
229 (vi) ORIGINAL SOURCE:
230 (A) ORGANISM: Helicobacter pylori
231
232 (ix) FEATURE:
233 (A) NAME/KEY: misc_feature
234 (B) LOCATION: 1...195
235 (D) OTHER INFORMATION: /note= "FLAGELLAR HOOK-ASSOCIATED PROTEIN 1
236 HAP1"
237
238 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:
239
240 GTGGAAAACA ACAAGAGTTT AAAGCATGCG AATGAGTTAA GGGATAAGCG AGATGAATTA 60
241
242 GAGTTTCATT TGCGAGAGCT TTTCGGGGGG AATGTTTTA AAAGCAGCAT TAAAACCCAT 120
243
244 TCGCTCACAG ATAAAGACTC AGCGGACTTT GATGAGAGCT ATAACCTTAA TATCGGGCAT 180
245
246 GGGYTCAATA TSATA 195
247
248
249 (2) INFORMATION FOR SEQ ID NO:6:
250
251 (i) SEQUENCE CHARACTERISTICS:
252 (A) LENGTH: 1857 base pairs
253 (B) TYPE: nucleic acid
254 (C) STRANDEDNESS: double
255 (D) TOPOLOGY: circular
256
257 (ii) MOLECULE TYPE: DNA (genomic)
258

DDDDDD CCCCCC SSSSS 22222
DD DDD CC C SS SS 22 22
DD DD CC SSS 22
DD DD CC SSS 222
DD DD CC SSS 222
DD DDD CC C SS SS 22 22
DDDDDD CCCCCC SSSSS 2222222

DDDDDD CCCCCC SSSSS 22222
DD DDD CC C SS SS 22 22
DD DD CC SSS 22
DD DD CC SSS 222
DD DD CC SSS 222
DD DDD CC C SS SS 22 22
DDDDDD CCCCCC SSSSS 2222222

Username DCS_2
Machine DCS_2
Comment

Date 97.03.25
Time 10:10:48
Printer @PRINTER
Device LPT1
Width 80
Tabs 0
Copies 1
File \spool.net\38._SP

PAGE: 6

**RAW SEQUENCE LISTING
PATENT APPLICATION US/08/487,632A**

DATE: 03/25/97

TIME: 10:10:36

INPUT SET: S16452.raw

******* PREVIOUSLY ERRORED SEQUENCES - EDITED *******

DDDDDD CCCCCC SSSSS 22222
DD DDD CC C SS SS 22 22
DD DD CC SSS 22
DD DD CC SSS 222
DD DD CC SSS 222
DD DDD CC C SS SS 22 22
DDDDDD CCCCCC SSSSS 2222222

DDDDDD CCCCCC SSSSS 22222
DD DDD CC C SS SS 22 22
DD DD CC SSS 22
DD DD CC SSS 222
DD DD CC SSS 222
DD DDD CC C SS SS 22 22
DDDDDD CCCCCC SSSSS 2222222

Username DCS_2
Machine DCS_2
Comment

Date 97.03.25
Time 10:11:03
Printer @PRINTER
Device LPT1
Width 80
Tabs 0
Copies 1
File \spool.net\38._SP

40867 (2) INFORMATION FOR SEQ ID NO:880:
40868
40869 (i) SEQUENCE CHARACTERISTICS:
40870 (A) LENGTH: 363 amino acids
40871 (B) TYPE: amino acid
40872 (D) TOPOLOGY: linear
40873
40874 (ii) MOLECULE TYPE: protein
40875
40876 (iii) HYPOTHETICAL: YES
40877
40878 (vi) ORIGINAL SOURCE:
40879 (A) ORGANISM: Helicobacter pylori
40880
40881 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:880:
40882
40883 Met Lys Phe Phe Leu Leu Lys Lys Phe Ser Xaa Phe Leu Asn Thr Gln
40884 1 5 10 15
40885
40886 Thr His Phe Asn Leu Lys Arg Leu Asn Ala Ser Ser Phe Leu Leu Glu
40887 20 25 30
40888
40889 Thr Phe Ser Lys Glu Lys His Ala Phe Val Val Asp Leu Ser Ala Pro
40890 35 40 45
40891
40892 Tyr Ile Gly Leu Ser Lys Lys Pro Pro Glu Ser Val Leu Lys Asn Thr
40893 50 55 60
40894
40895 Leu Ala Leu Asp Phe Cys Leu Asn Lys Phe Thr Lys Asn Ala Lys Ile
40896 65 70 75 80
40897
40898 Leu Gln Ala Asn Val Ile Asp Asn Asp Arg Ile Leu Glu Ile Lys Gly
40899 85 90 95
40900
40901 Ala Lys Asp Leu Ala Tyr Lys Ser Glu Thr Phe Ile Leu Arg Leu Glu
40902 100 105 110
40903
40904 Met Ile Pro Lys Lys Ala Asn Leu Met Ile Leu Asp Gln Glu Lys Cys
40905 115 120 125
40906
40907 Val Ile Glu Ala Phe Arg Phe Asn Asp Arg Val Ala Lys Asn Asp Ile
40908 130 135 140
40909
40910 Leu Gly Ala Leu Pro Pro Asn Ile Tyr Glu His Gln Glu Glu Asp Leu
40911 145 150 155 160
40912
40913 Asp Phe Lys Gly Leu Leu Asp Ile Leu Glu Lys Asp Phe Leu Ser Tyr

RAW SEQUENCE LISTING
PATENT APPLICATION US/08/487,032A

DATE: 03/25/97

TIME: 10:11:04

INPUT SET: S16452.raw

40914	165	170	175
40915			
40916	Gln His Lys Glu Leu Glu His Lys Lys Asn Gln Ile Ile Lys Arg Leu		
40917	180	185	190
40918			
40919	Asn Ala Gln Lys Glu Arg Leu Lys Glu Lys Leu Glu Lys Leu Glu Asp		
40920	195	200	205
40921			
40922	Pro Lys Thr Leu Gln Leu Glu Ala Lys Glu Leu Gln Thr Gln Ala Ser		
40923	210	215	220
40924			
40925	Leu Leu Leu Thr Tyr Gln His Leu Ile Asn Arg Arg Glu Asn Arg Val		
40926	225	230	235
40927			240
40928	Ile Leu Lys Asp Phe Glu Asp Lys Glu Cys Met Ile Glu Ile Asp Lys		
40929	245	250	255
40930			
40931	Ser Met Pro Leu Asn Ala Phe Ile Asn Lys Lys Phe Thr Leu Ser Lys		
40932	260	265	270
40933			
40934	Lys Lys Lys Gln Lys Ser Gln Phe Leu Tyr Leu Glu Glu Asn Leu		
40935	275	280	285
40936			
40937	Lys Glu Lys Ile Ala Phe Lys Glu Asn Gln Ile Asn Tyr Val Arg Asp		
40938	290	295	300
40939			
40940	Ala Ala Glu Glu Ser Val Leu Glu Met Phe Met Pro Val Lys Asn Ser		
40941	305	310	315
40942			320
40943	Lys Ile Lys Arg Pro Met Asn Gly Tyr Glu Val Leu Tyr Tyr Lys Asp		
40944	325	330	335
40945			
40946	Xaa Lys Xaa Gly Leu Gly Lys Thr Lys Lys Arg Ile Ser Ser Phe Tyr		
40947	340	345	350
40948			
40949	Lys Thr Gln Xaa Arg Met Ile Xaa Gly Cys Xaa		
40950	355	360	
40951			
40952			
40953			
40954			

PAGE: 1

SEQUENCE VERIFICATION REPORT
PATENT APPLICATION US/08/487,032A

DATE: 03/25/97

TIME: 10:11:07

INPUT SET: S16452.raw

Line

Error

Original Text